

Documento de Trabajo 96-04
Serie de Estadística y Econometría 02
Febrero 1996

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

EL FUTURO DE LOS METODOS ESTADISTICOS

Daniel Peña*

Resumen

Este trabajo analiza como la creciente accesibilidad inmediata a grandes masas de datos, la mayor potencia con menor coste de los métodos de cálculo, el crecimiento de la demanda por las técnicas estadísticas y la emergente necesidad de combinar distintos tipos de información van a impulsar el avance de nuevos métodos estadísticos. Se concluye que es previsible un desarrollo importante de técnicas exploratorias multivariantes, de métodos de estimación más generales y flexibles, de una mejor evaluación de la incertidumbre al construir modelos estadísticos, de métodos de predicción multivariante y de metodologías para Meta-Análisis. Se estudian algunas implicaciones de estos avances en la estadística oficial, y en especial en la actividad de un Instituto Nacional de Estadística (INE), resaltando como pueden contribuir a mejorar la calidad de la información suministrada por el INE y su servicio a la sociedad española.

Palabras clave

Computación estadística; Estadística oficial; Meta análisis.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.

1. INTRODUCCIÓN

En 1965 John W. Tukey, uno de los estadísticos más grandes de su época, predecía que la revolución informática produciría cambios en las técnicas estadísticas en cuatro direcciones principales (Tukey, 1965):

- a) sustitución de las tablas de las distribuciones por programas informáticos;
- b) mayor énfasis en los métodos basados en el orden de las observaciones;
- c) métodos de Monte Carlo de experimentación con el ordenador;
- d) procedimientos más eficientes de realizar los cálculos algebraicos.

Treinta años después podemos apreciar que los cambios en las técnicas estadísticas se han realizado en otras direcciones de las previstas por Tukey. De los cuatro aspectos, sólo los métodos de Monte Carlo han tenido un impacto fundamental en la evolución de las herramientas y técnicas estadísticas en estos 30 años, mientras que las direcciones a) y d) han tenido escasa influencia. Por el contrario, las técnicas de análisis de series temporales, tanto con modelos ARIMA como con modelos en el espacio de los estados, los métodos de estimación recursiva, los métodos de estimación robustos y no paramétricos o el análisis de datos discretos y categóricos mediante modelos lineales generalizados y loglineales han experimentado un crecimiento extraordinario y se han convertido en herramientas habituales del análisis estadístico.

Desde el punto de vista de la metodología, Tukey tuvo más éxito: predijo modelos más flexibles y un auge de los procedimientos gráficos y exploratorios, que efectivamente se ha producido. Estas predicciones metodológicas han sido fundamentalmente correctas, ya que Tukey también intuyó el cambio que estaba experimentando la metodología estadística al pasar de la secuencia lineal: modelo-datos muestrales-estimación-contraste de hipótesis al proceso cíclico de reformulación: modelo-datos-inferencia-diagnosís-modelo, donde el modelo se modifica a partir de los datos en sucesivas iteraciones de un proceso continuo de aprendizaje.

Este ejemplo señala que es más fácil prever las tendencias metodológicas generales que los avances en técnicas específicas. También ilustra las dificultades de intentar predecir la evolución de una disciplina como la estadística que al encontrarse en la frontera del avance de numerosas ciencias como 'la tecnología del método científico' - en la conocida definición de Mood - depende decisivamente de las demandas que recibe de sus campos de aplicación, que impulsarán la generación de nuevos métodos y técnicas.

Sin embargo, prever la evolución de una disciplina científica es importante por varias razones:

- a) orienta sobre el tipo de avances a los que debemos prestar una atención especial;
- b) sugiere que líneas de investigación son potencialmente más prometedoras;
- c) contribuye a la comprensión de las posibilidades y limitaciones de los métodos existentes.

Este trabajo presenta una visión personal, y por lo tanto forzosamente limitada por los sesgos y experiencias del autor, sobre la previsible evolución futura de los métodos estadísticos. El trabajo está estructurado de la forma siguiente. En la sección 2 se comentan algunos de los

cambios tecnológicos y sociales que pueden influir de manera destacada sobre la evolución de la Estadística. En la sección 3 se analizan las áreas de desarrollo más prometedoras que van a venir impulsadas por los cambios anteriores y que son potencialmente generadoras de nuevas técnicas estadísticas. Finalmente, la sección 4 presenta algunos comentarios sobre el efecto de estos desarrollos en la estadística oficial, con especial atención a la actividad del Instituto Nacional de Estadística.

2. CAMBIOS EN EL ENTORNO

Existe considerable evidencia de que el desarrollo de los métodos estadísticos ha venido condicionado por las demandas en otras disciplinas científicas y por los métodos de cálculo existentes. Este aspecto aparece claramente ilustrado con numerosos ejemplos en los libros de Box (1978) y Stigler (1986). Peña (1983) ha descrito la influencia de las teorías de Newton y Darwin en la evolución de la estadística matemática.

Extrapolando esta experiencia histórica, es razonable prever que la evolución futura de nuestra disciplina va a depender, como en el pasado, de las demandas de otras disciplinas, del tipo de información existente para el análisis estadístico y de las posibilidades tecnológicas de cálculo. Como tendencias futuras en estas direcciones resaltaremos únicamente las siguientes.

1. Accesibilidad inmediata a enormes bases de datos.

La utilización masiva del ordenador como instrumento de recogida de datos en procesos industriales y comerciales pondrá a disposición del usuario grandes masas de datos dinámicos y multivariantes. La recogida electrónica de información por lectores ópticos y sensores permite obtener con un coste marginal despreciable una enorme cantidad de datos socioeconómicos. Estas grandes bases de datos estarán cada vez más disponible para su análisis estadístico.

El INE comenzó en 1983, bajo la Dirección de Javier Ruiz-Castillo, la distribución de la Encuesta de Presupuestos Familiares a grupos de investigación universitarios. Desde esta experiencia pionera, en la actualidad es posible acceder a varias de las grandes encuestas que se realizan en España, tanto por el INE como el CIS y otros organismos de la Administración. Por otra parte, el usuario de Internet puede disponer de series financieras, con frecuencia de un segundo (el tiempo necesario de teclear las cifras en el ordenador) de la cotización de las principales bolsas internacionales, así como a un amplio conjunto de grandes bancos de datos estadísticos.

Estas tendencias se acelerarán en el futuro: por un lado, el avance de los ordenadores facilitará la recogida de información, generando la proliferación de enormes bases de datos. Por otro, el avance en la telecomunicaciones, la democratización política y la evolución hacia una sociedad cada vez más abierta y transparente hará que estos datos estén disponibles de forma inmediata y con un coste de acceso despreciable a un número creciente de usuarios potenciales.

2. Ordenadores más potentes y veloces.

La tendencia de disminución de los costes del hardware es previsible que continúe todavía en el futuro, unida a un aumento de la potencia de cálculo de los ordenadores personales y de su capacidad de manejo gráfico de la información. Hace 15 años el coste de un ordenador personal (tipo AT, con 640K y un disco de 20 Megs) era superior al precio de un automóvil de gama media. Hoy un ordenador personal corriente cuesta el 10% del precio de un automóvil y su capacidad y rapidez se han multiplicado por un factor de al menos diez. En términos reales su precio se ha dividido por un factor de cien.

La evolución de los sistemas operativos y programas de ordenador (software) ha sido paralela a la evolución de los equipos (hardware). En particular los programas para el cálculo estadístico han evolucionado de su orientación inicial al trabajo por lotes, donde se pedía al ordenador que realizase una tarea concreta y finalizase la ejecución, (como las versiones antiguas de BMDP o SPSS) a programas interactivos donde el usuario puede aplicar distintos análisis a un mismo conjunto de datos, con acceso directo a los resultados intermedios y capacidad de programación dentro del paquete (como en SAS, SCA, MATLAB, GAUSS etc) y a lenguajes orientados a objetos que permiten manejar indistintamente funciones, variables o gráficos (como el lenguaje S-plus, y algunas versiones recientes de los paquetes clásicos). En esta evolución los programas estadísticos han aumentado mucho su potencia y su flexibilidad, y además se han simplificado. Hoy existen paquetes estadísticos muy potentes que cualquier usuario puede utilizar con enorme facilidad y sin necesitar conocimientos avanzados de estadística.

Estos cambios, que continuarán previsiblemente en el futuro, posibilitarán la utilización masiva por los usuarios en campos muy diversos de técnicas más potentes y complejas para el análisis de la información disponible. Es previsible que en muy pocos años y en cualquier lugar del mundo un estudiante de estadística elemental tenga a su alcance, en su ordenador personal, una potencia de cálculo mayor de la disponible en los mejores centros de investigación estadística del mundo hace solo unos pocos años.

3. Mayor demanda de las técnicas estadísticas.

La existencia de una cantidad creciente de información cuantitativa producirá un crecimiento paralelo de la demanda por técnicas para el análisis de estos datos, es decir, de las técnicas estadísticas. Este fenómeno es ya visible en las revistas de investigación en cualquier campo científico (una proporción enorme de artículos en revistas de Medicina, Economía o Sociología utiliza técnicas estadísticas para justificar los resultados), pero tendrá una tendencia creciente como consecuencia del aumento de la formación estadística de los graduados universitarios en todas las ciencias. La estadística aparece ya en el Bachillerato y como disciplina básica en la gran mayoría de las titulaciones superiores y medias, tanto en las Ciencias Sociales (Sociología, Psicología, Ciencias de la Documentación, Economía etc) como en las Ingenierías, las ciencias sanitarias y biológicas y las Ciencias del medio ambiente. Esta posición singular va a potenciar la aplicación masiva de la estadística a nuevos problemas y la difusión de los avances en una area a las restantes áreas de aplicación.

La demanda por las técnicas estadísticas está también creciendo en el mundo de la empresa y la gestión de las organizaciones ligado al creciente interés por los temas de calidad. En un mundo cada vez más competitivo e interrelacionado, la calidad de los productos y servicios se está convirtiendo en un factor clave de supervivencia para las organizaciones que se mueven en entornos de libre competencia. Es sintomático que tanto el premio Deming a la calidad, que se concede anualmente a la empresa japonesa más destacada en la mejora de la calidad de sus productos y servicios, como el premio Malcolm Baldrige, para las empresas de EE.UU, incluyen, entre otros criterios a valorar, la utilización de herramientas estadísticas para el control y la mejora de los procesos en la empresa.

4. Necesidad de métodos para combinar distintos tipos de información

La difusión de los métodos estadísticos y la proliferación de estudios sobre el mismo problema, pero realizados con datos y condiciones experimentales distintas, está creando la necesidad de diseñar nuevos métodos para sintetizar y resumir esta información heterogénea. Los métodos estadísticos actuales suponen la existencia de una muestra homogénea, y son poco orientativos de como unificar distintos datos tomados bajo distintas condiciones y precisión experimental. Los métodos bayesianos son, en principio, más flexibles para combinar información subjetiva y objetiva, pero también requieren desarrollos importantes para incorporar de manera eficiente las variadas fuentes de información que se presentan en la práctica.

3. CAMBIOS EN LAS TÉCNICAS Y LOS MÉTODOS ESTADÍSTICOS

Las tendencias apuntadas en la sección anterior pueden inducir avances importantes en la metodología y las técnicas estadísticas para dar respuesta a las demandas que se plantean. Sin entrar en herramientas específicas, cuya evolución resulta mucho más incierta, son esperables desarrollo metodológicos especialmente relevantes en las siguientes áreas:

1. Técnicas exploratorias multivariantes.

La existencia creciente de grandes bancos de acceso general creará una fuerte demanda por técnicas exploratorias para su análisis, ya que el primer problema es sintetizar y comprender la información disponible en ellos. Actualmente estamos empezando a utilizar las enormes posibilidades del ordenador para la exploración de bases de datos multivariantes. Por ejemplo, un camino prometedor para el análisis exploratorio y gráfico es proyectar los puntos sobre ciertas direcciones elegidas para revelar determinados aspectos de su estructura. Estas técnicas, cuyo antecedente es el análisis de componentes principales, se conocen con el nombre de 'Projection Pursuit'. Su objetivo es la búsqueda de representaciones de los datos en espacios de dimensión reducida que se definen maximizando una función de utilidad determinada. Su desarrollo está siendo posible por la capacidad y rapidez de los ordenadores actuales. La investigación en este área puede abrir nuevas perspectivas en muchos problemas de análisis multivariante como métodos de clasificación de observaciones, de reducción de dimensionalidad, de reconocimiento de formas y patrones etc. Una revisión de las técnicas

de Projection Pursuit se encuentra en Jones y Sibson (1987).

Los avances en técnicas gráficas de análisis de datos van a tener también repercusiones importantes en la forma en que se construyan los modelos estadísticos. El reciente libro de Cook y Weisberg (1994) es un brillante ejemplo de las nuevas perspectivas que el análisis gráfico puede ofrecer para la construcción de modelos de regresión.

Una segunda línea de avance serán los desarrollos en métodos no paramétricos de exploración de grandes bases de datos. La utilización de estos métodos esta sujeta a la 'maldición de la dimensionalidad' (el aumento exponencial de las necesidades de datos o de tiempo de cálculo al aumentar la dimensión del espacio de las variables) y su aplicación en varias dimensiones presenta en la actualidad limitaciones muy importantes. Sin embargo, son excelentes en pocas dimensiones para identificar la estructura subyacente, por lo que pueden utilizarse para analizar ciertas proyecciones relevantes de los datos, aportando una gran flexibilidad en su análisis.

2. Métodos de estimación más generales y flexibles.

Los métodos clásicos de estimación estadística, como el método de máxima verosimilitud, requieren disponer de una muestra homogénea de una distribución que se supone conocida salvo por un número de parámetros. Estos métodos se encuentran con dos problemas principales:

- a) no son aplicables cuando el modelo es desconocido;
- b) son muy poco robustos a desviaciones de las hipótesis del modelo.

En consecuencia, son necesarios nuevos métodos para tratar el tipo de información compleja y heterogénea con que nos encontraremos al analizar grandes masas de datos. Estos desarrollos parecen ir en dos direcciones principales:

- a) Nuevos métodos de estimación que puedan aplicarse en condiciones muy generales. El ejemplo más claro son las técnicas de remuestreo (bootstrap), donde el ordenador realiza automáticamente el proceso de inferencia y proporciona directamente los intervalos de confianza o los errores estándar para el estimador deseado. Estas técnicas están proporcionando ya respuestas simples, gracias al uso extensivo del ordenador, en muchos problemas complejos (Efron y Tibshirani, 1993). Sin embargo, su utilización con muestras dependientes, como en el análisis de series temporales, esta todavía por desarrollar.
- b) Nuevos métodos de estimación adaptados al objetivo preciso que se pretende con el modelo. Un ejemplo de estos posibles desarrollos son las técnicas de previsión adaptativa (Tiao y Xu, 1993) donde la estimación de los parámetros se hace minimizando el error de predicción al horizonte considerado y cambia con este horizonte.

3. Mejor evaluación de la incertidumbre del proceso de inferencia.

La selección del modelo más adecuado, ha sido un área importante de investigación estadística en los últimos años, y disponemos de un conjunto de criterios para la selección del mejor modelo de regresión o de series temporales ante una muestra concreta (véase Aznar, 1989 y O'Hagan, 1995). A continuación, la metodología clásica estadística procede suponiendo el modelo conocido y las únicas fuentes de incertidumbre consideradas son las debidas a la aleatoriedad de las variables, dados los parámetros del modelo, y la incertidumbre asociada a la estimación de estos parámetros. Sin embargo, la incertidumbre asociada a la selección del modelo es generalmente ignorada.

Algunos autores (Draper, 1995) han argumentado, desde un punto de vista bayesiano, que en lugar de seleccionar el modelo 'mejor' para los datos observados debemos asignar probabilidades a los modelos en litigio y utilizar este conjunto de modelos para la evaluación de la incertidumbre en nuestras predicciones. Recientemente han aparecido modelizaciones donde cada observación puede ser generada por un modelo distinto con cierta probabilidad (véase por ejemplo McCullogh y Tsay, 1994). Podemos decir que el paradigma clásico de un proceso único generador de los datos se está progresivamente transformando en un nuevo paradigma que considera toda una clase de procesos potencialmente generadores de los datos observados.

Estas formulaciones bayesianas han experimentado un impulso enorme con la aparición de nuevos métodos de computación basados en muestreo por Monte Carlo. Las posibilidades de cálculo existentes están produciendo una revolución al permitirnos trabajar con clases amplias de modelos que permiten una flexibilidad inimaginable en el pasado. Por otra parte, se está comprobando que los algoritmos existentes no siempre funcionan en problemas complejos (Justel y Peña, 1996), lo que está generando la necesidad de desarrollar nuevos algoritmos de computación más inteligentes y adaptativos que los actuales.

4. Predicción de series multivariantes.

En los últimos veinte años se ha avanzado mucho en la modelización y predicción de series univariantes. Sin embargo, la predicción de conjuntos de series temporales aprovechando la estructura de dependencia subyacente ha avanzado muy poco. Aunque existe una abundante literatura sobre la construcción de modelos multivariantes de series temporales su utilización práctica es todavía pequeña. Un problema a resolver es el crecimiento exponencial del número de parámetros con la dimensión del vector de variables. Una solución prometedora es condensar la información en un número de factores comunes, lo que reduce el problema a términos manejables. La existencia de grandes bancos de series temporales interrelacionadas (financieras, de tráfico, médicas, etc) va a estimular la necesidad de realizar investigación básica y aplicada sobre este importante problema.

5. Desarrollo de metodologías para Meta Análisis.

Cualquier estimador estadístico puede considerarse un método para combinar la información muestral respecto a un parámetro y cualquier método de predicción es también un

procedimiento para combinar la información existente en los datos respecto a la cantidad de interés en la predicción (Draper et al, 1992, Peña, 1995). Sin embargo, carecemos de métodos precisos para combinar la información proveniente de experiencias con distinto grado de precisión. Esto es lo que se denomina Meta-Análisis, que se define en la National Library of Medicine como: " Un metodo cuantitativo de combinar el resultado de estudios independientes y sintetizar las conclusiones" (Olkin, 1992). Su impacto en la estadística médica es espectacular (véase Olkin, 1992) pero es previsible su difusión en todas las áreas de aplicación de la estadística, lo que requerirá la creación de una teoría adecuada para estos objetivos.

Un problema de particular interés que puede englobarse dentro del esquema de Meta-Análisis es el problema de combinación de predicciones de distintas fuentes. (véase Draper et al, 1992, Draper, 1995 y Guerrero y Peña, 1995).

4. SU IMPACTO EN LA ESTADÍSTICA OFICIAL

Este trabajo ha sido elaborado para el 50 aniversario de la creación del Instituto Nacional de Estadística (INE) y parece pertinente preguntarse cómo estos cambios pueden afectar a la estadística oficial y en especial a la actividad del INE. Refiriéndonos a las cinco líneas presentadas en la sección anterior los efectos sobre las actividades del INE pueden ser los siguientes:

1. Nuevos métodos de control de calidad de la información.

Existe considerable evidencia de que la recogida masiva de información lleva asociada la aparición de observaciones atípicas provenientes de errores de medida o de problemas no detectados de homogeneidad. Estos errores se identifican fácilmente cuando producen el efecto de convertir en atípico el valor en una variable pero son muy difíciles de detectar cuando su efecto es sólo aparente al estudiar la relación conjunta entre varias variables. Por ejemplo, un error en el gasto en alimentación de una familia en la encuesta de presupuestos familiares que produce un valor enorme o muy pequeño de este gasto con relación a la distribución de esta variable será fácilmente detectado y corregido. Sin embargo, un error que conduce a un valor enorme del gasto solamente cuando comparamos esta familia con otras de igual composición familiar, nivel de educación y localización geográfica será mucho más difícil de detectar. Las técnicas de análisis de datos multivariantes discutidas anteriormente aportaran métodos para identificar estos valores atípicos, lo que permitirá corregirlos y mejorar el proceso de recogida de datos al identificar mejor las situaciones donde se producen estos errores. Este proceso redundará en estadísticas más fiables.

2. Utilización del análisis descriptivo en la presentación de los datos.

Existe cierta controversia sobre el papel relativo que los Institutos Nacionales de Estadística debe tener en el análisis y síntesis de la información que recogen (Véase por ejemplo Malinvaud, 1985 y Bjerve (1985)). Sin embargo, la mejora de la calidad de la información comentada en el punto anterior requiere el uso de técnicas de análisis. Por otro lado, es

previsible un fuerte aumento de la demanda social por información digerida y estructurada, lo que requerirá la utilización de técnicas descriptivas estadísticas para convertir los datos en información útil para el ciudadano y la Administración. Es obvio que hay una cierta distancia entre los datos brutos y la información necesaria para tomar decisiones racionales de administración pública. En el futuro es previsible que aumente la demanda para que el INE no solo proporcione datos brutos tabulados, sino que contribuya a convertir estos datos en información relevante para mejorar el conocimiento de la realidad económica y social de nuestro país.

3. Introducción de métodos de meta-análisis para reconciliar la información.

Las agencias estadísticas tienen que trabajar con distintas fuentes de información con respecto al mismo fenómeno y los datos tomados por distintas fuentes casi nunca coinciden exactamente. Un ejemplo de este hecho, que ha suscitado cierta polémica en España, es la medida del desempleo. Por otro lado, la información de flujos internacionales de bienes o personas debe ser coherente con la publicada en otros países. Por ejemplo, Malinvaud (1985) ha señalado que los datos que dos países presentan sobre sus flujos respectivos de turistas o comercio bilateral generalmente no coinciden. Las técnicas de Meta Análisis pueden proporcionar un marco objetivo para reconciliar estas discordancias y proporcionar una mejor información a los poderes públicos y a los ciudadanos.

Para la realización de esta tarea es imprescindible que cada base de datos lleve asociada la información relevante sobre su recogida, período de vigencia, fuentes de error, etc. Esto requerirá procedimientos estandarizados para la creación de estas bases de "meta datos".

4. Aumento de la diseminación de información

Es previsible un crecimiento exponencial de la demanda de universidades y centros de investigación públicos y privados por los datos estadísticos recogidos por el INE. Por otro lado, el camino iniciado por el INE de poner sus datos a disposición pública continuará en el futuro, merced al avance de las telecomunicaciones, permitiendo que los datos estén disponibles mediante Internet de manera instantánea. Esto ocurrirá no solamente en el INE, sino en otros Institutos europeos de Estadística y en Eurostat. La gran oferta de datos estadísticos oficiales generará una mayor demanda por una amplia gama de usuarios potenciales, que exigirá cambios en la organización y tratamiento de las bases de datos estadísticas oficiales. En particular los problemas de confidencialidad de la información y de seguridad de la misma tendrán cada vez mayor relevancia.

5. Medición estadística de nuevas realidades complejas.

Como ejemplo de estos fenómenos señalaremos que países como Suecia y EEUU han comenzado a publicar indicadores de calidad de los bienes y servicio nacionales. Estos indicadores son importantes porque el aumento del nivel de precios en un país tiene un efecto distinto si va unido a un aumento de la calidad de sus productos y servicio que si sucede en una situación de estancamiento o declive de los mismos. La medida de la calidad presenta retos nuevos para los institutos de Estadística y requiere el desarrollo y asimilación de modelos complejos de análisis estadístico. Es previsible que la Unión Europea decida la

implantación generalizada de estos indicadores de calidad lo que supondrá un importante reto para las agencias estadísticas oficiales.

5.CONCLUSIONES

La metodología estadística ha venido muy condicionada por las posibilidades de cálculo existentes. Si consideramos las hipótesis que realizamos en un problema concreto como restricciones debidas a las limitaciones de los datos y de nuestros medios de cálculo, es fácil imaginar que muchas de las hipótesis de normalidad (y por tanto linealidad) independencia, homocedasticidad etc que realizamos al construir modelos tendrán una importancia menor en el futuro, en la medida en que dispongamos de más datos y de mejores medios de cálculo.

La posibilidad de calcular eficientemente ha permitido avances fundamentales en las técnicas y métodos estadísticos. Por ejemplo, la hipótesis de independencia se ha relajado considerablemente como consecuencia de los avances en el campo de series temporales. también se han hecho importantes avances en la comprensión de la no linealidad tanto en regresión (Seber y Wild, 1989) como en series temporales (Tong, 1990) y en otras muchas áreas. En el futuro, las posibilidades que proporcionarán los medios de cálculo para la exploración gráfica multivariante y la combinación de información es previsible que conduzcan a una metodología más flexible, más interactiva y más general de construcción de modelos estadísticos.

AGRADECIMIENTOS

Este trabajo ha sido financiado por la Cátedra BBV de Métodos para la mejora de la Calidad en la Universidad Carlos III de Madrid y por la DGICYT a través de proyectos PB93-0232 y PB94-0374. Agradezco a Felipe Aparicio, Costas Goutis, Francisco Melis, Jorge Muruzábal, Rosario Romera, Juan Romo y Santiago Velilla sus comentarios a una versión preliminar de este trabajo.

REFERENCIAS

- Aznar, A. (1989). Econometric Model Selection: A New Approach. Kluwer Academic Publishers.
- Bjerve, P. J. (1985). International Trends in Official Statistics. En A celebration of Statistics. Editado por A. C. Atkinson and S. E. Fienberg. Springer-Verlag.
- Box, J.F. (1978). R.A. Fisher. The life of a Scientist. Wiley: New York.
- Cook, R. D. y Weisberg, S. (1994). An Introduction to Regression Graphics. Wiley.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). The Journal of the Royal Statistical Society B, 57,1, 45-98.

Efron, B. y Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman and Hall.

Guerrero, V. y Peña, D. (1995). Linear Combining of Information in Time Series. Working Paper 95-52. Universidad Carlos III de Madrid.

Jones M.C. y Sibson R. (1987). What is Projection Pursuit?. Journal of the Royal Statistical Society B, 150, 1, 1-38.

Justel, A. y Peña, D. (1996). Gibbs Sampling Will Fail In Outlier Detection Problems With Strong Masking. The Journal of Computational and Graphical Statistics. August, 1996. (en prensa).

McCulloch, R. E. y Tsay, R. S. (1994). Bayesian Inference of Trend and Difference Stationarity. Econometric Theory, 10,596-608.

Malinvaud, E. (1985). Economic and Social Statistics for Comparative Assessments. En A celebration of Statistics. Editado por A. C. Atkinson and S. E. Fienberg. Springer-Verlag.

O'Hagan A. (1995). Fractional Bayes Factors for model comparison (with discussion). The Journal of the Royal Statistical Society B, 57,1, 99-138.

Olkin, I. (1992). Meta-Analysis: Methods for Combining Independent Studies. Statistical Science, 7,2,226.

Peña, D. (1983). La influencia de Newton y Darwin en la evolución de la Estadística Matemática, Estadística Española, 99, pp.: 103-121. 1983.

Peña, D. (1995). Combining Information in Statistical Modelling. Working Paper 95-51. Universidad Carlos III de Madrid.

Seber, G. A. F. y Wild, C. J. (1989). Nonlinear Regression. Wiley

Stigler, S. M. (1986). The History of Statistics. Harvard University Press.

Tiao, G. C. y Xu, D. (1993). Robustness of maximum likelihood estimates for multi-step predictions: The exponential smoothing case. Biometrika, 80,3,623-41.

Tong, H. (1990). Non-linear Time Series. Oxford Science Publications.

Tukey J. W. (1965). The Technical Tools of Statistics. The American Statistician, 19, 2, 23-28.